[1]道德外语效应及其调节变量：来自元分析的证据*

朱麟 [1] 刘瑾茹 [1] 李静 [1] 刘聪慧 [1]

[1] 中国人民大学心理学系，北京 100872

摘 要 同一个道德决策情景使用外语 (相比母语) 呈现时，个体会表现出更强的功利性倾向，即道德外语效应。随着研究的深入，结论并不一致。本研究运用元分析方法首次探讨了语言类型 (母语 vs. 外语)对道德判断中功利性倾向的影响，并分析了相关的调节变量。通过文献检索及梳理，共有 19 篇文献 46 个独立样本 97 个效应量符合元分析标准 (N = 9672)。结果显示存在较小但稳定的道德外语效应 (g = 0.23)；调节效应分析表明，道德外语效应受故事类型的影响，在个人道德两难故事中存在较小但稳定的外语效应 (g = 0.32)，但在非个人道德两难故事 (g = 0.11) 与日常道德评价故事中 (g = 0.12) 不存在外语效应；非个人道德两难故事中的外语效应受记分方式的影响，多点记分在该故事类型下存在效应 (g = 0.27)，二点记分不存在效应 (g = 0.05)；性别和语系类型没有显著的调节效应。这些结果表明语言类型对个体面对道德困境时的选择倾向有一定程度的影响，道德故事类型和记分方式在未来的研究中需要加以考虑。

关键词 道德判断，外语效应，元分析，调节作用

分类号 B842.1

1 引言

当个体面临风险决策，需要在两个描述损失的选项中做出选择时，会倾向于风险寻求，选择风险较大的选项；当个体需要在两个描述收益的选项中做出选择时，会倾向于风险规避，选择风险较小的选项，此现象被称为框架效应(Tversky & Kahneman, 1981)；但是，当同样的问题用外语呈现时，相较于母语条件，被试的选择并未呈现出框架效应，即个体的决策偏差降低，这一现象被称为外语效应(关旭等, 2017; 张诗容, 胡平, 2017; Keysar et al., 2012)。之后，有研究者在道德判断领域也发现了该效应(Costa et al., 2014)。当我们面临道德两难困境，例如需要杀死 1 个“胖子”拯救 5 个人时，个体更倾向于无所作为；但当同一困境也用外语呈现时，相较于母语条件，被试更倾向于杀死“胖子”挽救 5 个人，即个体决策结果更具有功利性倾向。外语效应在风险决策领域与道德判断领域的表现有所不同：风险决策领域中所

收稿日期：2020-11-09

*中国人民大学科学研究基金（中央高校基本科研业务费专项资金资助）项目成果(20XNA028)资助。

通信作者：刘聪慧, E-mail: liuconghui2001@ruc.edu.cn

呈现的外语材料让个体更倾向于做出理性的判断；而道德判断领域中所呈现的外语材料让个体更倾向于做出功利性的判断，有研究专门提出了将道德困境中所展示的外语效应称为“道德外语效应”(Cipolletti et al., 2016)。

需要强调的是，在这一系列有关道德外语效应的研究中，个体习得外语的年龄较晚，12岁左右，习得形式主要在课堂教学环境 (高园园, 2017; 张诗容, 胡平, 2017; Pavlenko, 2012)。此外，在本元分析中的 46 个独立样本中，有 17 个样本的被试，相较于本国，外语国家具有更高的社会经济发展水平(如汉语与英语，印度语与英语)；而剩下 29 个样本的被试，本国和外语国家具有相似的社会经济发展水平(如法语与德语，法语与英语)。

自“道德外语效应”被发现以来(Costa et al., 2014)，很多研究者发现了显著的效应(林蓝菲, 高华, 2020; Cipolletti et al., 2016; Corey et al., 2017; Geipel et al., 2015a, 2015b, 2016; Hadjichristidis et al., 2015; Hayakawa et al., 2017)，但也有相当数量的实验没有重复出外语效应(Bialek et al., 2019; Brouwer, 2019; Cavar & Tytus, 2017; Driver, 2020; Dylman & Champoux-Larsson, 2020)，即该效应是否稳定存在争议，是否有重要的调节变量还需要进一步的探讨(Hayakawa et al., 2016; Polonioli, 2018)。元分析方法可以整合多项有关道德判断领域中外语效应的实证研究成果，以有效避免单一研究结果中存在的抽样及测量误差(Thacker, 1988)；同时，对大量实证结果的系统回顾和调节变量检验，有利于准确把握能调节道德外语效应的重要变量(Higgins & Thompson, 2002; Polonioli, 2018)。本研究首次使用元分析的方法考察道德外语效应的强度，也进一步探究了哪些因素可以调节该效应的强弱，对回答现阶段该领域存在的争议、指导未来实证研究具有启发意义。

本文首先对道德外语效应的内部机制和稳定性进行了介绍，并梳理了可能调节该效应的变量；其次，对近 10 年的相关文献进行了元分析，包括异质性、发表偏差检验，以及主效应和和调节效应分析；最后，结合已有证据对主效应和调节效应进行了深入讨论，并对未来研究进行了展望。

2 道德外语效应

在道德判断任务中，语言类型会影响被试的判断结果。当天桥版本的电车困境用外语呈现时，相比于母语，被试更倾向于做出功利性选择，即牺牲 1 个人拯救 5 个人(Costa et al., 2014)。但道德两难故事更多是反映个体哲学范畴上的道德倾向，为使道德判断更加贴近现实情境，研究者将日常生活的不道德行为作为判断材料，发现在外语条件下，被试对不道德行为的宽容度更高(Geipel et al., 2015b)。在此基础上，同批研究者将道德判断的实验材料进

行了更加细致的划分，分为结果积极但动机不良的故事和结果糟糕但动机积极的故事；结果发现，被试在外语条件下更注重行为结果而忽略动机本身(Geipel et al., 2016)，表明外语条件下，被试更倾向于做出注重结果的功利性选择。综上，在道德判断领域，决策结果常与“伤害或牺牲无辜者”及“不道德行为的宽容程度”有关，而外语条件下的道德决策多反映个体对于最终结果的看重，即更具功利性倾向，表现为拯救大多数人或更宽容的道德评判。这种语言类型对道德判断中功利性倾向的影响被称为道德外语效应(Cipolletti et al., 2016, 张诗容, 胡平, 2017)。

2.1 道德外语效应的内部机制

道德是人类社会在长期的历史进程中逐渐形成且为大部分社会成员接受和遵守的行为准则(Graham et al., 2011)。道德判断是指个体以道德准则为标准对事件或行为进行善恶好坏、是非曲直的判别与划分(喻丰等, 2011)。而影响道德判断心理机制的因素到底是认知还是情绪，则是心理学领域中经久不衰的主题；也对应着道德心理学中的道德推理与道德直觉(邱俊杰, 张锋, 2015)。道德推理和道德直觉均代表着一种道德判断的心理过程，前者是基于既定的知识对道德情境进行推理的认知过程(Harenski et al., 2010)，而后者是根据道德情境的情绪反应直接进行的、快速的直觉过程(余宏波, 刘桂珍, 2006)。但是，道德判断究竟是认知推理的产物，还是情绪启发式的结果一直是众多学者争夺的焦点，各种理论模型被相继提出，具有代表性的包括：认知发展观、启发式理论、社会直觉模型和双加工模型(喻丰等, 2011)。认知发展观和启发式理论均认为道德判断是一种认知过程，与情绪无关；但认知发展观强调道德判断是道德推理的过程，道德直觉在其中无特定作用(Crain, 1985)；而启发式理论则指出道德直觉在道德判断中占主要地位，道德推理的作用仅体现在事后解释(Baron, 1994; Sinnott-Armstrong et al., 2010)。随着心理学领域情绪研究的兴起，情绪对道德判断的影响逐渐受到学者关注，社会直觉模型便应运而生；该理论强调道德判断依旧是依靠道德直觉，但道德直觉是由情绪所驱动，因此道德判断体现得是个体在面临道德情境时的情绪过程，而以道德推理为主的认知过程仅反映在事后解释中(Haidt, 2001, 2007, 2008)。但是 Cushman 等人(2010)认为既需要强调情绪在道德判断中的重要地位，也不能否定基于审慎推理的认知过程在其中的作用。由此，Cushman 等人(2010)提出了道德判断的“双加工模型”，并强调情绪与认知在道德判断中均存在无可替代的作用。

目前有关道德外语效应内在机制的理论解释主要集中于“双加工模型”，即分析式的理性加工和直觉式的情绪加工(高园园, 2017; Cushman et al., 2010; Kahneman, 2003)。分析式的理性加工基于道德推理过程，依赖于理性分析和逻辑推理，需要进行一定程度的认知努力，

因此会耗费更多的时间进行决策，使个体更倾向于功利论的判断(喻丰等, 2011)；而直觉式的情绪加工基于道德直觉过程，依赖于启发式的、情绪驱动及经验联结的直觉过程(田学红等, 2011)；该加工过程不需要投入过多的认知资源、凭借丰富的经验与自动化的技能快速做出判断，使个体更倾向于道义论的判断(Kahneman & Frederick, 2002)。研究还发现，情绪是影响个体进行道德决策的重要心理机制(胡艺馨等, 2018; Moll & de Oliveira-Souza, 2007)，当道德决策过程由直觉式加工主导时，个体可能会唤起更强烈的情绪反应；反之，当分析式加工在道德决策中占主导时，被试唤起的情绪反应较弱。而外语对个体情绪反应的减弱作用，在外语效应提出前就有较为充足的证据(Caldwell-Harris, 2014; Harris et al., 2003; Pavlenko, 2004; Puntoni et al., 2009)。对被试生理反应的观测同样得到了一致的结论(Harris et al., 2003, 2006)。基于道德判断的“双加工”理论，结合语言类型对情绪的影响，对道德外语效应的内部机制有如下解释：分析式加工理论认为，阅读外语需要调用个体额外的认知资源且增加个体认知负担，进而降低了个体面对道德两难困境的情绪唤醒强度，让信息加工的过程变得更加谨慎，从而导致被试更注重“最多救几个”的功利论取向(张诗容, 胡平, 2017; Costa et al., 2014)。进一步研究发现，相较于母语，外语呈现的同一事件不仅会增加认知载荷，使得个体认知加工的流畅性降低(Segalowitz, 2010)，还会降低个体对于信息的直觉反应(Geipel et al., 2016)。而直觉式加工理论主张，母语情境下，个体更多依赖情绪驱动的直觉启发式加工(罗跃嘉等, 2013)；这一加工是自动化、无需认知努力、经验联结的过程，其中情绪唤醒起着重要的作用(胡艺馨等, 2018; Pereira & Praxedes, 2016)；反之，外语情境下个体的情绪唤醒水平更低，基于情绪唤醒的道义论取向在加工信息时，其自动化过程被弱化，进而更倾向于功利性的选择结果(关旭等, 2017; Geipel et al., 2015a)。

2.2 道德外语效应是否稳定存在？

研究者们通过“双加工”理论来解释出现了道德外语效应的实验结果：如主张分析式加工的 Cipolletti 等人(2016)认为语言类型的改变增加了个体认知负担，导致其更依赖于分析式加工。但支持直觉式加工的 Hadjichristidis 等人(2015)的研究指出，双语条件下的决策过程并不包含向分析式加工的转换，而是弱化了情绪驱动的启发反应。其实“双加工”理论并非相互对立的，更多是反应个体不同的加工倾向，从而导致对于道德两难故事的决策差异(关旭等, 2017; 尹军等, 2018)。而这种对于道德判断的加工倾向是受语言类型影响的：外语条件下，倾向于分析式加工的个体会更加主动的调动认知资源理解故事、权衡不同结果的利弊，更加审慎的处理故事信息，进而更关注结果最大化的选择，即功利性选项(尹军等, 2018; Costa et al., 2014; Geipel et al., 2015b)；而倾向于直觉式加工的个体，在理解外语呈现的故事时难

以自动化的处理信息，情绪反应被弱化，对于故事中受害者的共情程度更低，为此更少选择道义论选项(Geipel et al., 2015a; Hayakawa et al., 2017)。基于以上论述，出现了道德外语效应的研究中，不论个体倾向于何种信息加工方式，面对外语呈现的道德决策情境时，其信息加工的结果是趋于一致的，即外语条件下，个体在道德判断中更倾向于功利性的决策结果。而这种倾向的稳定性很可能基于双语条件下，个体情绪唤醒水平存在差异的稳定证据。而在未重复出道德外语效应的实验中，研究者们也发现了影响该效应稳健性的其它因素，这些因素一般也与个体情绪唤醒有关：如绝大部分研究只在个人道德两难故事中发现了语言类型的影响，而不同类型的故事具有不同的自我卷入度，进而导致情绪反应的差异(Costa et al., 2014; Geipel et al., 2015a, 2016; Shin & Kim, 2017)。也有研究发现，当个体外语的熟练程度较高，也难以观察到外语效应，这可能是由于双语情绪反应的差异随语言熟练度的提升而减弱(Wong & Ng, 2018)。还有研究者从语言文化相似性的角度切入，发现双语间文化相似性越高的群体越不存在语言类型的效应，对此可能的解释是文化相似性会影响双语情绪唤醒水平的差异，使不同语言条件下被试的情绪反应相近，进而导致道德决策结果的倾向性没有差异(Cavar & Tytus, 2017; Dylman & Champoux-Larsson, 2020)。综上，道德外语效应的稳健性与双语条件下个体情绪反应的差异息息相关，但由于该效应也受其它因素的干扰，语言类型的影响程度可能较为有限(倪传斌, 2020)；所以，我们既要重视但又不能夸大语言类型的单独影响。由此，本研究提出假设 1：语言类型对道德判断中的功利性倾向存在一定程度的稳定影响。

2.3 道德外语效应的调节变量

自道德外语效应被发现以来，研究者们开展了广泛的有关其稳健性及影响因素的相关研究，但是研究结果不尽相同，仍存在相当数量的部分研究没有重复出道德外语效应，这让研究者对其稳健性提出质疑。基于文献梳理，我们认为道德外语效应的稳定性可能与选用的故事类型、研究对象的性别、结果的记分方式及双语间的语系类型等因素有关。

2.3.1 故事类型

故事类型可能调节道德外语效应。有研究者发现，道德外语效应并非稳定出现在所有的道德两难故事中，天桥版本的电车困境(是否亲自推下一个人)相较于转换版本的电车困境(是否变换轨道)更容易观察到语言类型的影响(Corey et al., 2017; Geipel et al., 2015a)。以上两种版本的电车困境根据决策者施加伤害行为的个人卷入程度存在一个明确的划分，即个人道德两难故事与非个人道德两难故事(Moore et al., 2011)。这两种故事类型的区别在于，决策者是否对受害者施加了直接的伤害行为，如天桥困境亲自推人阻止电车便是个人道德两难困境，

而电车困境仅仅转动开关阻止电车便是非个人道德两难困境。个人道德两难故事强调行为的主动性与直接性、更易唤醒情绪反应，非个人道德两难困境强调行为的间接性、情绪唤醒水平更低，而外语可以显著减弱个体的情绪反应(Dewaele, 2004; Pavlenko, 2004)。由此推断，个人道德两难故事相较于非个人道德两难故事，其情绪唤醒程度更易受语言类型的影响，进而更易出现决策结果的偏差。研究结果也表明，一般在个人道德两难困境中观察到的外语效应较为稳定，因为该情境被试的自我卷入程度更高，在母语条件下更容易引发强烈的冲突情绪，而外语对于情绪的弱化作用，导致不同语言条件下情绪唤醒的差异，进而造成决策结果的不同(Geipel et al., 2016; Hadjichristidis et al., 2017; Hayakawa et al., 2017)。在非个人道德两难故事中被试自我卷入程度不高，由此不论何种语言条件下，被试的情绪反应差异不大。综上，本研究提出假设 2：故事类型能够对道德外语效应起调节作用。

2.3.2 性别

性别可能调节道德外语效应。首先，从共情的性别差异分析：生理层面上，催产素有助于情绪共情，而睾酮则与认知共情有负相关，两性的共情水平也由此存在差异；性别角色层面上，当两性习得各自的性别角色以后，其在情绪共情方面的差异较为突出(陈武英等, 2014)。因此，面对同一道德两难困境，女性相较于男性对故事中的受害者可能存在更高水平的共情，诱发更多的负性情绪，进而对决策结果产生影响。其次，从情绪调节来看，女性对于负性情绪的调控效果不如男性，男性更善于利用抑制情绪的方式对负性情绪进行日常管理(蔡阿燕等, 2016)。此外，道德规范敏感性的性别差异可能来源于睾酮分泌水平的性别差异(曾笑雨, 马燚娜, 2020)。因此，相较于男性，女性面对相同的道德两难困境时，对负性情绪进行抑制调控的难度更大，进而更倾向于道义论选择。综上，本研究提出假设 3：性别能够对道德外语效应起调节作用。

2.3.3 记分方式

结果的记分方式也可能会对道德外语效应产生影响。对本次纳入元分析的文献进行梳理可知，研究者使用的结果记分方式存在差别，主要存在二点记分(“是”选项或“否”选项)与多点记分(询问牺牲无辜者的意愿程度)两大类，研究者通过不同的记分方式，获得的效应值大小依旧存在差异(Geipel et al., 2015a)。例如，Geipel 等人(2015a)的研究结果显示，在天桥版本的电车困境中利用二点记分获得的结果刚刚达到显著的标准，其等价的效应量 d 为 0.32，而相同的实验程序利用多点记分获得的等价效应量 d 为 0.95。根据效应值 d 的划分标准(Cohen, 1992)，二点记分获得的语言效应属于较小的效应值，而多点记分获得的结果属于较大的效应值，多点记分可能比二点记分更加敏感、更易获得较大的效应量。这可能是由于

二点记分方式的选项呈现较为单一，更关注个体对于道德两难困境的结果而不重视对某种决策结果的倾向性程度。而多点记分方式相较于二点记分方式更加敏感，可能更容易捕捉到被试在不同语言条件下面对同一困境的某种倾向性(Geipel et al., 2015a)。因此，本研究提出假设 4：记分方式能够对道德外语效应起调节作用。

2.3.4 语系类型

双语间的语系类型也可能会调节道德外语效应。语言学家们将同源性程度较高、一般结构相似性较高的不同语言纳入同一语系(Holman et al., 2011)。通过对纳入该元分析的文献梳理可知，被试的母语与外语一般可以归为：隶属同一语系或不同语系。同一语系下的语言具有更高的结构相似性，而当母语与外语的结构相似性较高时，外语可以通过激活母语的一般结构，进而获得唤醒水平较为一致的情绪体验(Green, 2009)。反之，隶属不同语系的语言间不具有较高相似性的语言结构，个体在阅读外语呈现的道德困境时，无法依托母语与外语共享的语言结构或语法进行故事内容的理解或情绪体验迁移(Dylman & Barry, 2018)，更多依靠个体本身的外语水平，由此导致不同语言条件下的情绪差异。有研究者发现，当母语与外语间的语系类型及文化交流密切相关时，即使在个人道德两难故事中也未观察到语言效应(Brouwer, 2019; Cavar & Tytus, 2017; Dylman & Champoux-Larsson, 2020)；反之，外语效应在隶属不同语系、使用群体相距更远的语言组合中更易出现(Brouwer, 2019)。由此推论，同一语系下的语言由于其语言结构的相似性，母语与外语间的情绪唤醒水平类似，进而双语的决策结果更可能趋于一致；反之，不同语系下的语言更易出现情绪差异，可能更容易观察到外语效应。综上，本研究提出假设 5：语系类型能够对道德外语效应起调节作用。

3 研究方法

3.1 文献搜集

2012 年，“外语效应”在风险决策领域中被首次提出(Keysar et al., 2012)，随后研究者将该效应逐渐拓展到道德判断领域(Costa et al., 2014)，并不断验证其稳定性及影响因素(Dylman & Champoux-Larsson, 2020)，由于该效应从提出至今还不到十年，为此本研究将检索时间跨度设为 2010 年 1 月到 2020 年 5 月，涵盖近十年的有关研究，以中文和英文文献检索方式进行文献检索。依据已见刊的中英文文献中题目、摘要、关键词这三部分的高频词作为文献检索的关键词。按照顺序依次在中国知网、万方、维普数据库将关键词“外语效应”、“道德外语效应”、“外语”、“双语”分别与“道德”、“道德判断”、“道德决策”、“道德评价”、“道德

两难故事”进行匹配检索，搜索篇名、关键词及摘要包含此类关键词的期刊与论文。接着将关键词：“foreign language effect”、“foreign language”、“bilingual”、“moral foreign language effect”、“the effect of foreign language”分别与“morality”、“morals”、“moral decision-making”、“moral judgement”、“moral dilemma stories”、“moral evaluation”，在 Web of Science 核心合集、Pub Med、Science Direct、Springer Online Journals 数据库中分别进行检索，搜索篇名、关键词及摘要中包含此类关键词的文献。未发表的中文文献通过中国优秀硕士学位论文全文数据库、博士学位论文全文数据库进行查找；未发表的英文文献通过邮件询问该领域的主要研究者和查找 ProQuest 和 DART-Europe 硕博论文数据库进行搜集。

3.2 文献纳入与排除标准

将检索的文献按照以下标准进行筛选：(1)研究必须是考察道德判断中外语效应的实证研究，数据资料完整明确，排除纯理论或综述类文章；(2)研究对象为正常人群，其它特殊群体(例如存在各类心境障碍的被试群体)需要被排除；(3)文章需考察道德判断中的外语效应，并明确报告了能转换成 d 值的 χ^2 值、F 值或 t 值，若文章未报告以上参数则向作者进一步讨要，讨要未果再排除； (4)考虑研究结果异质性的问题，若研究中同时存在生理指标与行为指标，则应排除生理指标，保留行为指标；(5)排除重复发表的文章，同一批数据反复使用的应只统计一次。最终纳入元分析的文献共计 19 篇，包括中文文献 3 篇、英文文献 16 篇，共计 97 个效应量，文献检索及筛选流程图见图 1。

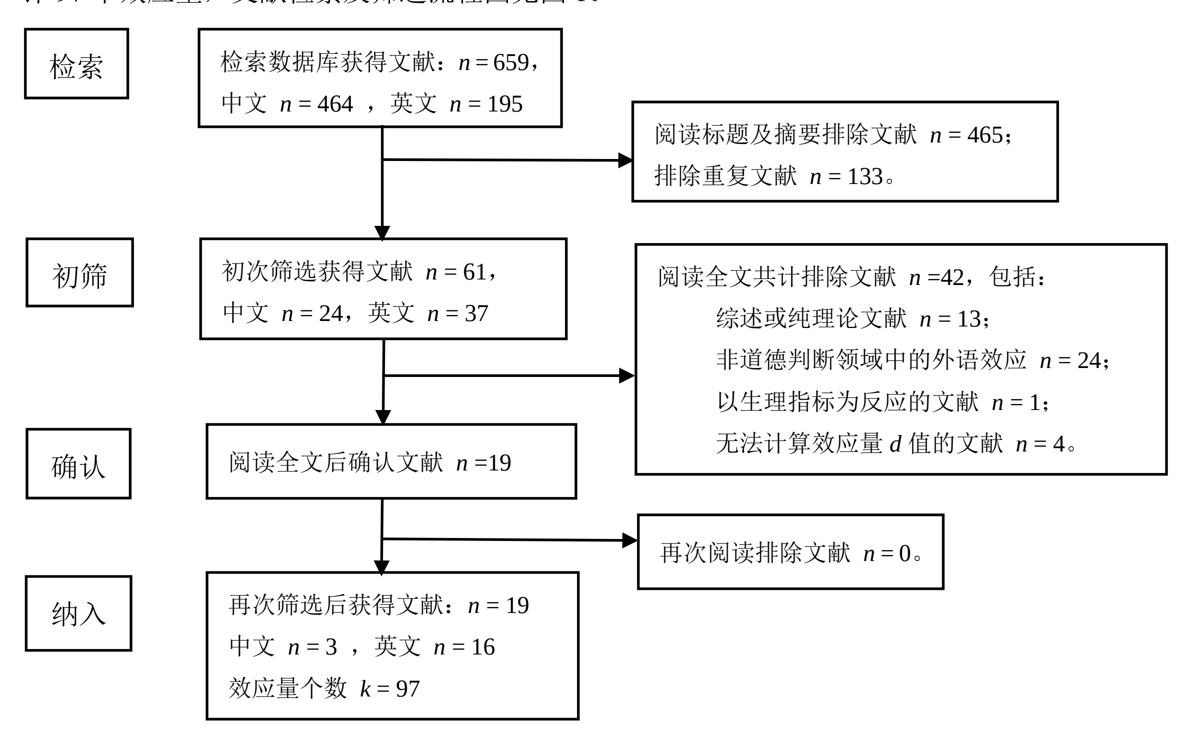


图 1 文献筛选流程图

3.3 文献编码

对纳入元分析的文献进行如下编码：(1)文献信息(作者及发表年限)；(2)实验顺序；(3)该实验采用的故事数目；(4)纳入分析的故事次序/数目；(5)样本量；(6)女性占比(%)；(7)故事类型；(8)记分方式；(9)语系类型；(10)效应值 g。效应值的提取遵照以下标准：(1)由于一部分研究者在分析数据时，会依次分析每个故事的效应，一部分研究者会直接分析所有故事的效应；因此，每项研究中独立的故事编码一次，若存在多个独立故事则分开编码；直接分析所有故事的样本编码一次。(2)同一批数据重复发表的文章只取其一，以期刊论文为主。该研究的文献编码工作由两位编码者独立提取、同时进行(Borenstein et al., 2011)，两位编码者均为中国人民大学心理学系的硕士研究生，具有较强的文献阅读及统记分析能力，两位编码者的一致性为 96.11%，说明该研究的文献编码比较准确且有效。此外，编码不一致的研究也通过进一步讨论而达成共识。最终纳入元分析的文献共计 19 篇，包括中文文献 3 篇、英文文献 16 篇，共计 97 个效应量(表 1)。

考虑到本研究纳入的文献大部分设置了多项实验、逐次分析了每个故事的效应，进而得到的多个效果量会占据较大的权重，可能会导致结果产生偏差(Borenstein et al., 2011)；同时传统的元分析统计方法无法处理非独立效应量之间的相关性，为此该研究将使用可处理效应量间相关的 Robust variance estimation (RVE)统计方法，并与传统的元分析结果进行对比，进而了解多个非独立效应量之间的相关性对该研究结果的影响 (Fisher & Tipton, 2015)。

表 1 纳入分析的原始研究的基本资料

文献	实验顺序	道德故事数目	纳入分析的故事次序/数目	样本量	女性占比(%)	故事类型	记分方式	语系类型	效应值 g
杨雯雯, 2020	Exp.1	4	1/1	44	61.37	日常道德评价故事	多点记分	不同语系	0.048
			2/1	44	61.37	日常道德评价故事	多点记分	不同语系	0.008
			3/1	44	61.37	日常道德评价故事	多点记分	不同语系	-0.117
			4/1	44	61.37	日常道德评价故事	多点记分	不同语系	-0.289
胡宏安, 2019	Exp.1	8	1-8/8	74	50.00	非个人道德两难故事	多点记分	不同语系	0.199
	Exp.2	8	1-8/8	70	50.00	个人道德两难故事	多点记分	不同语系	1.181
关旭, 2017	Exp.1	8	1-8/8	62	100.00	个人道德两难故事	二点记分	不同语系	1.126
	Exp.2	8	1-8/8	62	100.00	个人道德两难故事	二点记分	不同语系	-0.043
	Exp.3	8	1-8/8	62	100.00	个人道德两难故事	二点记分	不同语系	-0.087
Hong Im Shin & Juyoung Kim, 2017	Exp.1	4	1/1	161	67.08	非个人道德两难故事	二点记分	不同语系	0.265
			2/1	161	67.08	非个人道德两难故事	二点记分	不同语系	0.166
			3/1	161	67.08	个人道德两难故事	二点记分	不同语系	-0.820
			4/1	161	67.08	个人道德两难故事	二点记分	不同语系	0.417
Alexandra S. Dylmana, & Marie-	Exp.2a	1	1/1	198	83.84	个人道德两难故事	二点记分	同一语系	-0.132
			1/1	175	84.00	个人道德两难故事	二点记分	同一语系	-0.451

France Champoux-Larsson, 2020	Exp.2b	1	1/1	305	67.87	个人道德两难故事	二点记分	同一语系	0.069
			1/1	295	78.21	个人道德两难故事	二点记分	同一语系	0.064
Janet Geipel et al., 2015	Exp.1	2	1/1	67	83.81	个人道德两难故事	二点记分	同一语系	0.557
			2/1	76	83.81	个人道德两难故事	二点记分	同一语系	0.716
		2	1/1	67	83.81	非个人道德两难故事	二点记分	同一语系	0.135
			2/1	76	83.81	非个人道德两难故事	二点记分	同一语系	0.401
	Exp.2	2	1/1	152	44.72	个人道德两难故事	二点记分	不同语系	0.330
			2/1	152	44.72	非个人道德两难故事	二点记分	不同语系	0.016
			1/1	152	44.72	个人道德两难故事	多点记分	不同语系	0.963
	Exp.3	4	1/1	72	76.39	个人道德两难故事	多点记分	同一语系	0.511
			2/1	72	76.39	非个人道德两难故事	多点记分	同一语系	0.152
			3/1	72	76.39	个人道德两难故事	多点记分	同一语系	0.484
			4/1	72	76.39	非个人道德两难故事	多点记分	同一语系	0.295
Susanne Brouwer, 2019	Exp.1	6	1/1	60	50.00	个人道德两难故事	二点记分	同一语系	0.196
			2/1	60	50.00	个人道德两难故事	二点记分	同一语系	0.196
			3/1	60	50.00	个人道德两难故事	二点记分	同一语系	-0.155
			4/1	60	50.00	非个人道德两难故事	二点记分	同一语系	-0.866
			5/1	60	50.00	非个人道德两难故事	二点记分	同一语系	1.516
			6/1	60	50.00	非个人道德两难故事	二点记分	同一语系	0.000
	Exp. 2	6	1/1	60	53.33	个人道德两难故事	二点记分	同一语系	0.654
			2/1	60	53.33	个人道德两难故事	二点记分	同一语系	0.165
			3/1	60	53.33	个人道德两难故事	二点记分	同一语系	0.541
			4/1	60	53.33	非个人道德两难故事	二点记分	同一语系	-0.355
			5/1	60	53.33	非个人道德两难故事	二点记分	同一语系	0.881
			6/1	60	53.33	非个人道德两难故事	二点记分	同一语系	0.695
Michele Miozzo et al., 2020	Exp.3	1	1/1	225	47.00	个人道德两难故事	二点记分	同一语系	0.290
	Exp.4	1	1/1	861	44.00	个人道德两难故事	二点记分	同一语系	0.368
Albert Costa et al., 2017	Exp.1	1	1/1	708	76.98	日常道德评价故事	多点记分	同一语系	0.035
	Exp.2	1	1/1	667	53.19	日常道德评价故事	多点记分	同一语系	0.011
Franziska Čavar & Agnieszka Ewa Tytus, 2017	Exp.1	6	1/1	60	51.67	个人道德两难故事	多点记分	同一语系	-0.165
			2/1	60	51.67	个人道德两难故事	多点记分	同一语系	0.003
			3/1	60	51.67	个人道德两难故事	多点记分	同一语系	-0.363
			4/1	60	51.67	个人道德两难故事	多点记分	同一语系	0.003
			5/1	60	51.67	个人道德两难故事	多点记分	同一语系	0.132
			6/1	60	51.67	个人道德两难故事	多点记分	同一语系	0.008
Meagan Y. Driver, 2020	Exp.1	2	1/1	180	74.64	非个人道德两难故事	二点记分	同一语系	-0.128
			2/1	180	74.64	个人道德两难故事	二点记分	同一语系	0.530
Evy Woumans et al., 2020	Exp.1	4	1/1	256	85.55	个人道德两难故事	多点记分	同一语系	0.299
			1/1	256	87.11	非个人道德两难故事	多点记分	同一语系	0.281
			2/1	256	85.55	个人道德两难故事	多点记分	同一语系	0.531
			2/1	256	87.11	非个人道德两难故事	多点记分	同一语系	0.299
			3/1	256	85.55	个人道德两难故事	多点记分	同一语系	0.420
			3/1	256	87.11	非个人道德两难故事	多点记分	同一语系	0.461
			4/1	256	85.55	个人道德两难故事	多点记分	同一语系	0.373

			4/1	256	87.11	非个人道德两难故事	多点记分	同一语系	0.282
Rafal Muda et al.,	Exp.1	20	1-10/10	140	65.00	非个人道德两难故事	多点记分	同一语系	-0.073
2018			11-20/10	140	65.00	个人道德两难故事	多点记分	同一语系	0.901
Albert Costa et al.,	Exp.1	1	1/1	317	53.00	个人道德两难故事	二点记分	混合语系	0.298
2014	Exp.2	2	1/1	725	71.92	个人道德两难故事	二点记分	同一语系	0.585
			2/1	725	71.92	非个人道德两难故事	二点记分	同一语系	0.013
			1/1	328	73.00	个人道德两难故事	二点记分	同一语系	0.521
			1/1	397	71.00	个人道德两难故事	二点记分	同一语系	0.641
			2/1	397	71.00	非个人道德两难故事	二点记分	同一语系	-0.072
			2/1	328	73.00	非个人道德两难故事	二点记分	同一语系	-0.033
Joanna D. Corey et	Exp.1a	2	1/1	211	72.54	非个人道德两难故事	二点记分	同一语系	0.470
al., 2017			2/1	211	72.54	个人道德两难故事	二点记分	同一语系	0.324
	Exp.1b	2	1/1	173	84.00	非个人道德两难故事	二点记分	同一语系	-0.360
			2/1	173	84.00	个人道德两难故事	二点记分	同一语系	0.000
	Exp.2a	2	1/1	204	79.95	非个人道德两难故事	二点记分	同一语系	-0.092
			2/1	204	79.95	个人道德两难故事	二点记分	同一语系	-0.040
	Exp.2b	2	1/1	399	76.43	非个人道德两难故事	二点记分	同一语系	-0.095
			2/1	399	76.43	个人道德两难故事	二点记分	同一语系	-0.124
	Exp.3a	2	1/1	202	73.05	非个人道德两难故事	二点记分	同一语系	0.073
			2/1	202	73.05	个人道德两难故事	二点记分	同一语系	0.642
	Exp.3b	2	1/1	190	70.47	非个人道德两难故事	二点记分	同一语系	0.334
			2/1	190	70.47	个人道德两难故事	二点记分	同一语系	0.426
	Exp.3c	2	1/1	201	51.36	非个人道德两难故事	二点记分	同一语系	0.458
			2/1	201	51.36	个人道德两难故事	二点记分	同一语系	0.279
	Exp.3d	2	1/1	197	67.00	非个人道德两难故事	二点记分	同一语系	0.423
			2/1	197	67.00	个人道德两难故事	二点记分	同一语系	0.069
	Exp.3e	2	1/1	217	83.48	非个人道德两难故事	二点记分	同一语系	0.410
			2/1	217	83.48	个人道德两难故事	二点记分	同一语系	0.161
Heather Cipolletti et	Exp.1	2	1/1	160	76.25	非个人道德两难故事	二点记分	同一语系	-0.073
al., 2016			2/1	160	76.25	个人道德两难故事	二点记分	同一语系	0.692
Janet Geipel,	Exp.1a	4	1-4/4	48	70.83	日常道德评价故事	多点记分	同一语系	0.716
Constantinos	Exp.1b	4	1-4/4	64	87.50	日常道德评价故事	多点记分	同一语系	0.553
Hadjichristidis et al.,	Exp.2	4	1-2/4	78	78.21	日常道德评价故事	多点记分	同一语系	0.511
2015		4	3-4/4	78	78.21	日常道德评价故事	多点记分	同一语系	0.042
	Exp.3	4	1-4/4	74	90.54	日常道德评价故事	多点记分	同一语系	0.666
Janet Geipel et al.,	Exp.1	3	1-3/3	107	73.83	日常道德评价故事	多点记分	同一语系	0.168
2016	Exp.2	2	1-2/2	144	70.14	日常道德评价故事	多点记分	同一语系	-0.369
Yuen-Lai Chan et al.,	Exp.1	39	1-39/39	144	52.08	未区分故事类型	二点记分	不同语系	0.143
2016									

注：元分析文献作者在两人之内时均列出作者名，多于两人时只列出第一作者；当同年度同作者文献有两条时，加入第二作者以区分。

3.4 元分析过程

3.4.1 效应量计算

元分析以标准化均数差 Hedge's g，即 Cohen's d 的修正量作为道德判断中外语效应的效果量(Vøllestad et al., 2012)。Hedge's g 相较于 Cohen's d 可以提供更精确的估计，尤其是在小样本数据时；虽然两者的计算方式存在些许差异，但其大小划分是一致的(Goulden & Keith, 2006)。搜集文献中报告的效应值 d 后，通过以下公式进行转化：$g = d\,[1 - (3 / (4\,df - 1))]$，$df = n_1 + n_2 - 2$。若纳入的研究未报告效应值 d，则根据样本量、平均数和标准差等原始数据进行计算：$d = (M_1 - M_2) / s_{pooled}$，$s_{pooled} = [(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2 / n_1 + n_2 - 2]^{1/2}$。若纳入的研究未完整报告样本量、平均数、标准差等原始数据，则利用相应的公式通过原始数据的χ^2 值、F 值或 t 值进行效应量的转换：$d = 2\, [\, \chi^2 / (N - \chi^2)]^{1/2}$；$d = 2[F\,(n_1 + n_2) / n_1 n_2]^{1/2}$；$d = t\,(n_1 + n_2 / n_1 n_2)^{1/2}$。

3.4.2 模型的选定

元分析基本从固定效应模型(fixed effects model)或随机效应模型(random effects model)中选择一种进行综合效应的估计。固定效应模型预设研究间的差异只由随机误差导致，不同研究间的真实效应值是相同的；而随机效应模型假设不同研究间的真实效应值可以不一致，除了随机误差，研究群体、研究范式等因素也会导致研究间的差异(Borenstein et al., 2011)。本研究梳理文献发现，道德外语效应可能受性别比例、故事类型和记分方式等因素的调节，因而采用随机效应模型进行元分析。此外，利用异质性检验可以进一步验证模型选择的适切性，其中包括 Q 检验与 I^2 检验。Q 检验中若 $p < 0.05$ 则结果异质；I^2 检验对异质性程度进行划分，对高、中、低、无异质性的分界取值分别为 75%、50%与 25% (Higgins et al., 2003)。若 Q 检验结果不显著或 I^2 呈现低异质性，则考虑固定效应模型；反之，该研究更适合随机效应模型。

3.4.3 发表偏差

当已见刊的文献未能全面系统的代表该研究领域的全貌时，就说明出现了发表偏差(Rothstein et al., 2005)。发表偏差导致已见刊的文献中阳性结果远高于阴性结果，这将使元分析结果出现重大偏差，进而导致获得的效应值高于真实值(Kuppens et al., 2013)。针对发表偏差的问题，在文献搜集阶段便要尽可能检索未发表的文章；还应在之后的分析中，通过漏斗图(funnel plot)、Egger's 回归检验、Classic Fail-safe N 值(失安全系数)和 *p-curve* 技术来检验发表偏差。漏斗图是将各研究的效应量转化成可视化的散点图，是一种主观评价的方式，若不存在发表偏差，数据应左右对称分布、集中在中上部。Egger's 回归的结果不显著，则

认为不存在发表偏差(Egger et al., 1997)。Classic Fail-safe *N* 值(失安全系数)表示使研究失去统计学意义还需要纳入多少个研究，当该值大于 $5k + 10$(k 为效应量个数)时，说明不存在显著的发表偏差(Rothstein et al., 2005)。*p-curve* 检验中如果某研究是存在效应的，那么应该出现右偏态的 *p* 值分布，否则，可能存在发表偏差(Simonsohn et al., 2015)。

3.4.4 数据处理及分析程序

利用 R 语言的 metafor 和 robumeta 程序包进行元分析的异质性检验、发表偏差检验、主效应检验及调节效应检验(Fisher & Tipton, 2015; Schwarzer, 2007; Viechtbauer, 2010)，其中调节效应检验包括亚组分析与元回归分析。亚组分析将检验分类变量的调节作用，如：故事类型、记分方式和语系类型；元回归分析将检验连续变量的调节作用，如：女性占比。若调节效应不显著则使用 JASP 0.13.1 进行贝叶斯因子估计，分析该结果是否支持零假设。由于传统的 Meta 分析在固定效应模型或随机效应模型的前提下，利用 *Q* 检验计算各研究间的差异，该方法基于正态分布的假设，因此当难以符合正态分布、存在小样本资料时，传统方法分析受限，可能会遗漏重要信息(Smith et al., 1995)；而贝叶斯因子估计不受限于上述的经典统计学假设，可以根据样本信息获得后验分布情况，并在此基础上计算效应值和研究间的方差，相较于传统方法更加敏感，更不易遗漏重要信息(Warn et al., 2002)。因此，在完成传统的元分析流程后，对不显著的调节效应结果进行贝叶斯因子估计可能会获得有价值的额外信息(张亚利等, 2019)。为此该研究在调节效应结果不显著的情况下，使用贝叶斯因子估计对数据资料进行二次探索。

4 研究结果

4.1 异质性检验

将语言类型对道德判断中功利性倾向影响的元分析数据进行异质性检验，*Q* 检验的结果表明：*Q* 值达到显著水平，$Q_{(96)} = 344.79$，$p < 0.001$，说明本研究中纳入的各个效应量之间存在异质性。I^2 检验的结果表明，语言类型对道德判断中功利性倾向影响的 $I^2 = 72.16\%$，接近高异质性的分界取值 75% (Higgins et al., 2003)，表明观察变异中 72.16%是由效应值的真实差异造成的，27.84%是由随机误差造成的，表明研究间的变异存在组间误差干扰，各研究间存在较高异质性。综上，之后的元分析适合随机效应模型；以上分析结果也暗示，某些影响因素可能会造成研究间的差异，因此需要对其中的调节变量进行深入分析。

4.2 发表偏差检验

漏斗图(图 2)显示，效应值主要分布在漏斗图的上方，左右大致对称，该分布特点表明，有关道德外语效应的研究基本不存在发表偏差。Egger's 回归法的检验结果不显著，表明该研究纳入的效应量不存在严重的发表偏差，截距为 -0.08，95%CI 为[-1.04, 0.87]，$p = 0.862 > 0.05$ (Egger et al., 1997)。失安全系数值(Classic Fail-safe *N*)表示使研究失去统计学意义还需要纳入多少个研究，当该值大于 $5k + 10$(k 为效应量个数)时，说明不存在显著的发表偏差，本研究的失安全系数为 3994，远高于临界值 495 (Rothstein et al., 2005)。*p-curve* 检验结果(图 3)表明曲线呈右偏态分布，进一步分析发现结果显著的 42 个效应值具有证据价值 ($Z = -9.81$, $p < 0.001$)。以上结果表明，本研究不存在严重的发表偏差。

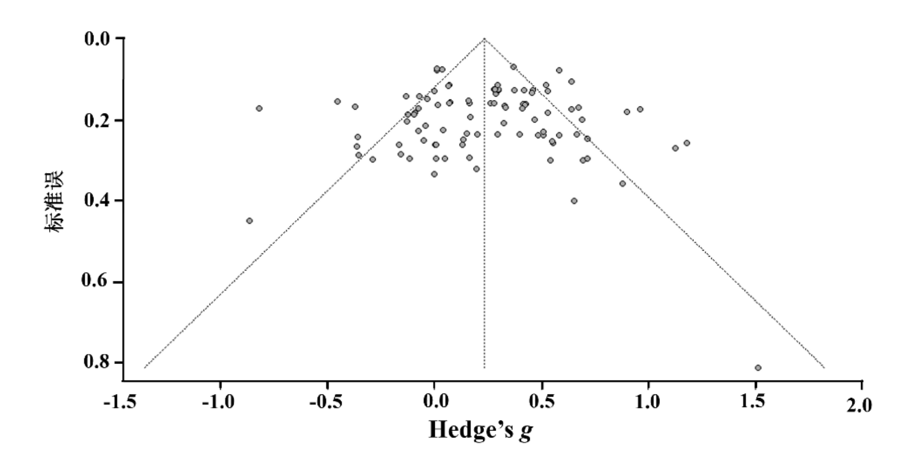
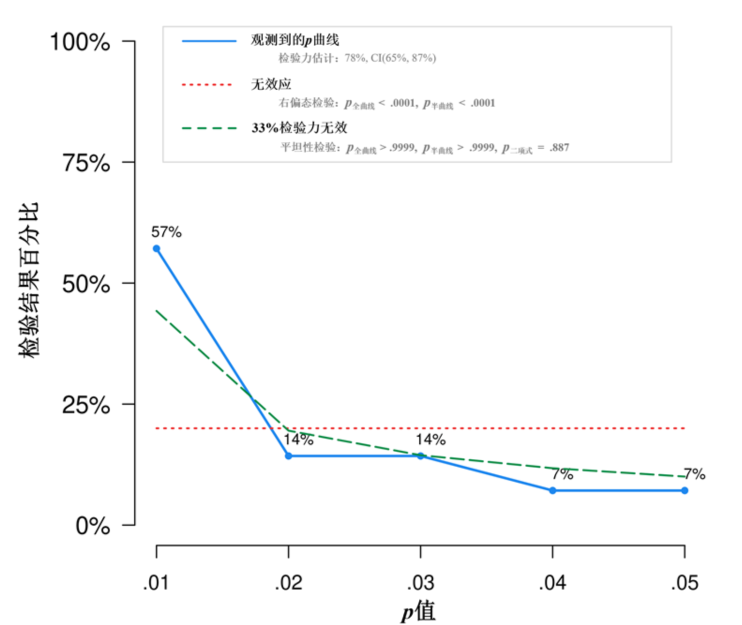


图 2 效应值分布漏斗图



注：观测到的 p 曲线包括42个统计上显著的结果($p < 0.05$)，其中33个 p 值低于0.025。有55个结果由于 p 值大于0.05没有参与 p 曲线检验。

图 3 *p-curve* 曲线图

4.3 主效应分析

采用随机效应模型将语言类型对道德判断中功利性倾向的影响进行综合性探讨，结果表明语言类型对道德判断中功利性倾向影响的主效应显著，$z = 6.75$，$p < 0.001$，效应量 g 为 0.23，95%CI 为[0.16, 0.29]。根据效应值 d 的划分标准，对大、中、小和无效应量的分界取值为 0.8、0.5、0.2 (Cohen, 1992)，则道德外语效应的效应值属于较小的效应量。对效应量进行敏感性分析，排除任意一个研究后的总效应量 g 值在[0.22, 0.24]间波动。根据森林图，同时剔除一项高异质性研究后(Susanne Brouwer, 2019-Exp.1-5/1)，效应量 $g = 0.22$，$z = 6.70$，$p < 0.001$。以上结果表明，该效应值较为稳定。

为避免某篇文献存在多个效应值、占据较大权重，导致结果偏差(Borenstein et al., 2011)；同时将基于随机效应模型的传统元分析统计方法与可处理非独立效应量间相关性的 Robust variance estimation (RVE)方法进行对比(Fisher & Tipton, 2015)，我们使用 robumeta 工具包对纳入该研究的 19 篇文献、共计 97 个效应量进行二次分析，结果表明：语言类型对道德判断中功利性倾向影响的主效应依旧显著，$t = 4.18$，$p < 0.001$，效应量 g 为 0.21，95%CI 为[0.10, 0.32]。该结果表明，即使纳入该研究的文献存在多个非独立的效应量，但是这些非独立效应量间的相关性对总效应量的影响是有限，最终的结果依旧是稳定的。

4.4 调节效应分析

对语言类型影响道德判断中功利性倾向的随机效应模型进行调节效应分析。结果表明，故事类型可调节道德外语效应；记分方式在非个人道德两难故事中的调节效应显著；性别和语系类型的调节效应均不显著。

故事类型能显著调节道德外语效应，亚组分析结果显著，Q 值(组间)为 10.96，$p = 0.004 < 0.05$；在个人道德两难故事中获得了较小但稳定的效应值 ($g = 0.32$)，但在日常道德评价故事 ($g = 0.12$) 与非个人道德两难故事中 ($g = 0.11$) 不存在道德外语效应。

性别的调节作用不显著。元回归分析表明，女性比例不能对道德外语效应的大小进行显著预测($b = 0.003$，95%CI 为[-0.002, 0.007]；$z = 1.06$，$p = 0.289$)。贝叶斯回归分析的结果显示，贝叶斯因子 BF_{10} 为 0.30，根据判定标准(Wagenmakers et al., 2017)，存在中等程度的证据认为道德外语效应不存在性别差异。根据上述分析，仅有个人道德两难故事存在稳定的道德外语效应，为此可进一步探究，性别能否调节个人道德两难故事中的外语效应。根据以往研究的建议，纳入元回归分析的研究不得少于 6 项(Fu et al., 2011)，使用个人道德两难故事的研究数目大于 6 项，所以该故事类型下的研究可进行元回归分析。结果表明：在使用个人

道德两难故事的研究中，元回归分析(50 项研究)结果不显著(b = 0.001，95%CI 为[-0.005, 0.007]；z = 0.34，p = 0.730)。

记分方式的调节作用不显著。亚组分析结果不显著，Q 值(组间)为 0.84，p = 0.358 > 0.05；但是其亚组结果显示出利用多点记分测得的效应值(g = 0.27)是高于二点记分的效应值(g = 0.20)。贝叶斯方差分析的结果也显示 BF_{10} 为 0.24，说明仅仅存在中等程度的证据认为道德外语效应不受记分方式的调节。根据上述分析，仅有故事类型能显著调节道德外语效应，为避免故事类型对结果的混淆，我们进一步探究了在不同故事类型中，记分方式的调节效应是否显著。结果表明：在使用个人道德两难故事的研究中，亚组分析(50 项研究)结果不显著，Q 值(组间)为 0.65，p = 0.421 > 0.05，多点记分效应值 g 为 0.38，二点记分效应值 g 为 0.29，均为较小但稳定的效应值；在使用非个人道德两难故事的研究中，亚组分析(33 项研究)结果显著，Q 值(组间)为 8.97，p = 0.003 < 0.05，多点记分效应值 g 为 0.27，存在较小但稳定的效应，二点记分效应值 g 为 0.05，不存在效应；使用日常道德评价故事的研究均为多点记分，无法进行亚组分析。根据以往研究的建议，亚组分析每组纳入的研究应不少于 4 项(Fu et al., 2011)，上述亚组分析结果符合该前提。

语系类型的调节作用不显著。亚组分析结果不显著，Q 值(组间)为 0.03，p = 0.874 > 0.05；同一语系效应值 g 为 0.23，不同语系效应值 g 为 0.21。贝叶斯方差分析的结果也显示 BF_{10} 为 0.28，说明存在中等的证据认为道德外语效应不受记分方式的调节。根据上述分析可知，依旧仅有故事类型能显著调节道德外语效应，为避免故事类型对结果的混淆，可进一步探究在不同故事类型中，语系类型的调节效应是否显著。结果表明，在使用任意一种故事类型的研究中，亚组分析结果均不显著：个人道德两难故事(49 项研究)，Q 值(组间)为 0.06，p = 0.815 > 0.05，同一语系效应值 g 为 0.32，效应量较小但显著，不同语系效应值 g 为 0.38，效应值不显著；非个人道德两难故事(33 项研究)，Q 值(组间)为 0.27，p = 0.604 > 0.05，同一语系效应值 g 为 0.11，不同语系效应值 g 为 0.16，均不存在效应；日常道德评价故事(13 项研究)，Q 值(组间)为 2.40，p = 0.121 > 0.05，同一语系效应值 g 为 0.19，不同语系效应值 g 为- 0.09，均不存在效应。根据以往研究的建议，亚组分析每组纳入的研究应不少于 4 项(Fu et al., 2011)，上述亚组分析结果符合该前提。亚组分析结果详见表 2；贝叶斯分析详见表 3；不同故事类型下，记分方式和语系类型的亚组分析结果详见表 4。

表 2 道德外语效应的调节效应检验

调节变量	异质性检验	类别	K	95%CI	双尾检验

	Q_B	df	p			点估计	下限	上限	Z	p
故事类型	10.956	2	0.004	个人道德两难故事	50	0.318	0.224	0.413	6.592	0.000
				日常道德评价故事	13	0.124	-0.034	0.281	1.536	0.125
				非个人道德两难故事	33	0.113	0.029	0.197	2.645	0.008
记分方式	0.790	1	0.373	二点记分	61	0.202	0.118	0.286	4.726	0.000
				多点记分	36	0.266	0.159	0.373	4.872	0.000
语系类型	0.030	1	0.874	同一语系	79	0.228	0.160	0.295	6.624	0.000
				不同语系	17	0.208	-0.028	0.443	1.731	0.083

注：Q_B 代表异质性检验结果；K 代表独立效果量的个数；95%CI 为亚组效果量 g 的 95%置信区间。

表 3 贝叶斯分析结果

	Models	P(M)	P(M|data)	BF_M	BF_{10}	%error
回归分析	Null model	0.500	0.771	3.376	1.000	
	性别	0.500	0.229	0.296	0.296	0.008
方差分析	Null model	0.200	0.622	6.577	1.000	
	记分方式	0.200	0.149	0.700	0.240	0.034
	语系类型	0.200	0.174	0.843	0.280	0.008

表 4 不同故事类型下，记分方式与语系类型的调节效应检验

调节变量	故事类型	异质性检验			类别	K	95%CI			双尾检验	
		Q_B	df	p			点估计	下限	上限	Z	p
记分方式	个人道德两难故事	0.646	1	0.421	二点记分	35	0.293	0.182	0.404	5.168	0.000
					多点记分	15	0.381	0.196	0.566	4.040	0.000
	非个人道德两难故事	8.967	1	0.003	二点记分	25	0.054	-0.042	0.149	1.104	0.270
					多点记分	8	0.272	0.166	0.378	5.026	0.000
语系类型	个人道德两难故事	0.060	1	0.802	同一语系	41	0.316	0.229	0.404	7.053	0.000
					不同语系	8	0.375	-0.108	0.859	1.521	0.128
	非个人道德两难故事	0.268	1	0.604	同一语系	29	0.106	0.011	0.201	2.175	0.030
					不同语系	4	0.157	-0.012	0.326	1.822	0.068
	日常道德评价故事	2.404	1	0.121	同一语系	9	0.188	-0.003	0.380	1.925	0.054
					不同语系	4	-0.087	-0.378	0.204	-0.587	0.557

注：Q_B 代表异质性检验结果；K 代表独立效果量的个数；95%CI 为亚组效果量 g 的 95%置信区间。

5 讨论

本研究利用元分析方法获得以下结果：(1)道德外语效应具有稳定但较小的效应量；(2)故事类型对该效应的调节作用显著；(3)记分方式仅在非个人道德两难故事中存在显著的调节作用；(4)性别和语系类型的调节作用均不显著。

5.1 语言类型对道德判断中功利性倾向的主效应

本研究将外语效应于 2012 年发现以来有关道德判断领域的 19 篇文献、共计 41 项研究进行了元分析，将语言类型对道德判断中功利性倾向的影响程度进行了分析与探索，结果发现外语的使用对道德判断存在较小但是稳定的效应量，即当个体使用外语而非母语进行道德决策时，更有可能做出功利性选择。该结果与前人的研究是一致(Cipolletti et al., 2016; Corey et al., 2017; Costa et al., 2014; Geipel et al., 2015b)，也对本研究的假设 1 进行了验证。

该研究发现，语言类型对道德判断中功利性倾向的影响(道德外语效应)是稳定的，这说明外语相较于母语确实在一定程度上会改变个体的道德决策倾向。前文论述的"双加工理论"的两类机制其实并非相互对立，更多是显示个体阅读外语时的某种加工倾向，导致道德判断偏向于功利性的决策结果，而双语导致的情绪反应差异在以上两条加工路径中均起着重要作用。分析式加工理论认为，外语情境下认知资源的调动与认知负担的增加降低了个体情绪唤醒强度，进而对信息处理更加审慎与理性(Costa et al., 2014)；反之，母语情境下，以耗费少量资源的自动化加工为主，更容易唤起个体情绪。而道德判断中功利性倾向与认知需求相关，道义性倾向则与情绪及共情有关(Conway & Gawronski, 2013)。直觉式加工理论主张，母语情境下社会道德规范的激活是依靠情绪为主导的直觉式加工，但外语情境弱化了这种以情绪驱动为主的启发式反应(Geipel et al., 2015a)，进而减弱了对社会道德规范的激活(Chan et al., 2016; Geipel et al., 2015a; Hadjichristidis et al., 2017)，从而使个体不易受其约束。由此可见，结合个体在双语间情绪反应差异的证据，"双加工理论"可以对道德外语效应的内在机制进行一定程度的解释。但其中难以忽视的客观情况是"个体在双语条件下接受了相同的道德情境信息，但却做出了不同的判断"，这说明提供给个体的信息即使其本质内容是相同的，但由于不同类型的语言载体，可能导致个体在关注和解释信息时存在差异，进而出现不一致的判断。

我们需要指出的是："双加工理论"并不能揭示信息以不同语言为载体呈现时，个体在其中进行信息加工的具体差别。而近年来兴起的解释水平理论则为语言影响道德判断的研究提供了新视角(李明晖, 饶俪琳, 2017)。解释水平理论认为个体对事物的信息表征可分为高低

两个解释水平，高解释水平的表征使个体更关注事物的核心特征和抽象性信息，而低解释水平的表征更强调事物的次要特征和细节化信息(Trope & Liberman, 2010)。心理距离是影响解释水平的重要变量，代表个体对某类事物以此时、此地、自己为参考点时接近或远离产生的一种主观经验(Liberman & Trope, 2014)；个体与某类事物的心理距离越远则进行表征的解释水平越高(Aguilar et al., 2013)。研究表明，人们在使用外语时对社会伦理的敏感性较低，伴随着更低水平的情绪唤醒(Dewaele, 2010)，进而增加了个体感知到的心理距离，由此针对外语呈现的信息更容易进行高解释水平，建立更加抽象的表征形式(张诗容, 胡平, 2017; Amit et al., 2009)；而母语的习得与情绪信息丰富的社会场景密切相关，使人们更在意道德规范与社会评价，更易诱发相关情绪反应，从而拉近了个体与相关事物的心理距离，对母语呈现的信息更易构建具体的、低解释水平的表征(Gong & Medin, 2012; Pavlenko, 2012)。此外，另有研究指出，当个体对道德判断进行高解释水平的抽象性表征时，更关注行为的结果，倾向于做出功利论的选择(Gong & Medin, 2012)；而对相关信息进行低解释水平的具体化表征时，个体更关注行为是否符合道德准则，进而倾向于道义论的判断(Aguilar et al., 2013)。由此推论，相同的道德情境用外语呈现时，个体可能更倾向于对其进行高解释水平的表征，进而更易出现功利论的判断；反之，相关信息以母语为载体时，由于低解释水平表征的可能性更大，道德判断倾向于道义论的结果。因此，未来对道德判断中外语效应的进一步探索可能需要聚焦于双语语境与解释水平间的关系及作用机制。

需要提及的是，本研究虽然聚焦于道德判断领域中的外语效应，且该效应在道德判断和风险决策领域的表现形式存在差异，但两者的内在机制也可能有相似之处。从最终结果来看，风险决策情境以外语呈现时，经典的决策偏差被削弱或消失，更符合理性加工的结果(张诗容, 胡平, 2017)；而在外语语境的道德判断中，个体所表现的功利性倾向暗示其决策时的加工过程更具理性思维(关旭等, 2017)。从两类决策包含的心理机制来看，道德判断是涉及认知控制、规则表征、情绪与推理、成本与收益评估等机制相互作用的复杂加工过程(李笑燃等, 2014)，而上述机制与风险决策密切相关(黄龙等, 2020)。此外，个体在进行道德判断时激活的神经网络与风险决策时调用的神经网络也有相似之处(罗跃嘉等, 2013)。这说明风险决策和道德判断虽然任务表现形式不同，但其内部加工过程有类似之处；而外语语境对道德判断和风险决策的影响，即外语效应的存在，可能与两种任务中情绪唤醒的程度有关(Geipel et al., 2015a)。相较于母语，外语对情绪唤醒的弱化及其离身性导致个体更倾向于分析式加工而非启发式加工(Costa et al., 2014, Geipel et al., 2016)。由于双语习得环境的差异，外语更具语义性信息而非情绪性信息(Pavlenko, 2004)，相同的决策场景通过双语表征后，其情绪性信

息和解释水平存在差异，使个体更关注最终的收益和结果(Shin & Kim, 2017)。因此，个体在外语语境下存在更多的理性决策行为或更加功利性的道德判断。

该研究还发现，外语对道德判断中功利性倾向的影响虽然稳定，但却属于较小的效应量，这表明语言是造成道德决策差异的重要因素而非决定性因素。这种选择倾向不仅受到语言条件的影响，还有其它因素会导致道德判断中的决策差异：如外语的文化影响力、母语与外语间的语言相似性会影响最终的决策结果(Dylman & Champoux-Larsson, 2020)；实验材料选用的故事类型也会影响外语效应的出现(Corey et al., 2017; Geipel et al., 2015a; Shin & Kim, 2017)；记分方式的不同使得同一批被试的实验结果其最终获得的效应值相差较大(Geipel et al., 2015a)。以上结果说明，尽管语言能对道德判断的结果产生某种倾向性的影响，但总体上看其影响程度依旧是有限的，远未达到预测个体决策结果的程度，因此，既要重视语言对于道德判断的作用，也不可过分夸大语言对决策结果的直接影响。该元分析结果也暗示了，需要利用调节效应分析进一步探索影响道德判断中外语效应出现边界的重要因素。

5.2 语言类型对道德判断中功利性倾向的调节效应

5.2.1 故事类型的调节效应

亚组分析结果显示，故事类型的调节效应显著，在个人道德两难故事中存在较小但稳定的道德外语效应，但非个人道德两难故事与在日常道德评价故事不存在道德外语效应。这表明故事类型会导致实验结果出现差异，支持了假设 3。采用不同的故事类型而出现的结果差异可能与被试面对道德情境的自我卷入程度有关。个人道德两难故事存在稳定的效应，可能是由于这样情境用母语呈现时让被试的自我卷入程度更高，从而会引发更多的负性情绪与冲突情感(张诗容, 胡平, 2017)。因此，被试以决策者的身份进行选择时，会拉近与受害者的情感距离，从而更倾向于做出道义性选择；而当该高情感冲突的情境用外语呈现时，由于外语对情绪的弱化作用，降低了被试的情绪反应，从而更倾向于功利性的选择。非个人道德两难故事与日常道德评价故事不存在效应，依旧与其诱发的道德情境的自我卷入程度有关。非个人道德两难故事的情境具有较低的自我卷入程度，被试在理解故事的基础上，不会因为语言的不同而产生情绪反应的差异，进而影响决策结果。且非个人道德两难故事获得的效应值大小处于无效应的范围内且结果不显著，这表明，相较于个人道德两难故事其难以存在稳定的外语效应，这与以往的研究一致(Geipel et al., 2015b; Hadjichristidis et al., 2017; Hayakawa et al., 2017)。在日常道德评价故事中也未发现道德外语效应，这是由于该类型的故事虽具有一定的真实性，但依旧不存在高情感冲突的情境，自我卷入程度不及个人道德两难故事。而使

用日常道德评价故事的研究中，仅有一半故事存在外语效应(Geipel et al., 2015b)，这说明该类型故事诱发的自我卷入程度也不尽相同。这一结果提示未来在确定道德判断的实验材料时，需要联系研究的具体目的进一步考虑故事诱发的自我卷入程度。

5.2.2 性别的调节效应

元回归分析结果显示，性别的调节效应不显著，贝叶斯回归分析也表明有中等程度的证据认为两者间的关系不受性别的影响。因此，该结果未能支持假设 2，表明道德外语效应可能存在跨性别的稳定性。虽然已有元分析的结果表明，对于道德困境问题的决策存在着性别差异，即相较于男性，女性更倾向于做出道义性选择(Friesdorf et al., 2015)，这可能是由于女性更容易唤起更强烈的负性情绪，具有更高的共情水平，而道义性倾向于共情有关(Conway & Gawronski, 2013)，这在一定程度上会使得女性更愿意做出道义性的选择。但是，本研究发现在外语条件下道德决策的结果不存在性别差异。这可能是因为，即使女性面对母语呈现的道德困境更容易唤醒更强烈的负面情绪，但是面对外语呈现的道德困境时，外语对被试情绪反应的弱化作用(Dewaele, 2004; Pavlenko, 2004)，导致负性情绪的唤醒水平在性别间的差异并不显著。这造成不同性别群体在外语条件下面对相同道德两难困境时具有相似水平的情绪状态，其道德两难德困境的决策结果便不一定存在性别差异。另外，贝叶斯分析表明仅有中等程度的证据支持不存在性别差异，未来可以进一步验证该结论的稳定性。

5.2.3 记分方式的调节效应

亚组分析还发现，记分方式的调节效应不显著。我们基于文献梳理发现，被试进行道德判断的方式有两大类，二点记分与多点记分。亚组分析结果表明，多点记分获得的语言类型对道德判断中功利性倾向的效应值要高于二点记分的效应值，但是两组间的组间差异并不显著，未能支持假设 4，但是却在一定程度上反映了假设 4 的基本趋势，即多点记分相较于二点记分更加敏感。贝叶斯方差分析的结果也显示，仅有中等程度的证据认为道德外语效应不受记分方式的调节。因此，两者间的关系是否不受记分方式的影响还需进一步的分析验证。出现该结果的原因可能在于，直接针对记分方式的亚组分析并未将故事类型这一因素进行区分；具有高自我卷入度的故事更容易观察到外语效应(Corey et al., 2017; Shin & Kim, 2017)而不受记分方式的影响(Geipel et al., 2015a)，而自我卷入度较低的故事可能需要依靠更加敏感的记分方式才能观察到外语效应。因此，在对采用的故事类型进行区分后，我们发现，在非个人道德两难故事中发现了记分方式的调节效应，而在个人道德两难故事中记分方式的调节效应则不显著。因此，在未来的研究成果更加丰富后可以进一步验证不同故事类型下，道德外语效应是否受到记分方式的调节。

5.2.4 语系类型的调节效应

亚组分析同样发现，语系类型的调节效应不显著，不同故事类型下也未发现语系类型的调节作用。我们根据文献提供的母语与外语的信息，将所有的研究分为：母语与外语属于同一语系或不同语系。亚组分析结果表明，组间差异不显著，未能支持假设 5。贝叶斯方差分析的结果显示，存在中等程度的证据认为道德外语效应不受语系类型的影响。出现该结果的原因可能在于，本次亚组分析有关不同语系的研究仅有 17 项，且这 17 项研究间本身就具有很高的异质性。已有研究表明，母语与外语间的相似性会对外语效应产生影响，语言相似的外语可以激活母语的一般结构，从而在用外语阅读道德困境时会激活体验水平较为一致的情感共鸣(Dylman & Champoux-Larsson, 2020)。这可能通过语言间共享的语言结构、共享的语法、同源词的数量等来实现。即使个体的外语水平较低，但是母语与外语具有较高的相似性，其在用外语阅读道德困境时，可能已经策略性地将外语锚定在他们母语的情感上(Colomé & Miozzo, 2010; Dylman & Barry, 2018; Green, 2009)。因此，我们推论属于同一语系的语言间可能更不容易观察到外语效应，而不同语系的语言间由于策略性的迁移过程难以实现可能更容易观察到外语效应，语系类型可能存在差异。由于相关的研究较少，可待拥有更丰富的研究成果后继续探究道德外语效应是否受到语系类型的调节。

综上，本研究首次利用元分析方法系统探究了外语效应在道德判断领域中的稳健性及影响因素，回答了研究者对道德外语效应稳定性的争议，发现了故事类型的调节作用，揭示了记分方式对效应值大小的潜在影响，对未来的实证研究在实验材料和统计方式的最终选择上具有一定的启示和指导意义。

5.3 研究不足与展望

首先，元分析需要竭尽周全的纳入已有的研究成果，该研究虽利用检索工具尽可能的对文献进行了搜集整理，但一些未发表的文献依旧难以搜集，存在数据遗留的问题。其次，外语效应的影响因素较多，但在本研究中仅仅考虑了性别、故事类型、记分方式和语系类型这 4 个因素，还有一些影响因素未被涉及到，如双语者外语的熟练程度越低，越容易观察到外语效应(Conrad et al., 2011; Degner et al., 2012)；由于本研究纳入的文献接近半数在实验中未搜集相关信息，数据量不足，因此未放入分析。最后，尽管本研究验证了同篇文献下各个非独立效应量间的相关对最终结果的影响有限，但一篇文献纳入多个效应量的情况在本研究中不可避免，将来的研究可以尝试建立元分析的三级整合模型，进一步区分研究内误差与研究间误差(Assink & Wibbelink, 2016)。未来研究应在更系统全面的搜集文献后，进一步拓展可调节道德外语效应的重要变量，并构建更加准确的数据分析模型来佐证研究假设。

6 结论

本研究通过元分析方法得出如下结论： (1)语言类型对道德判断中的功利性倾向存在稳定但有限的影响；(2)故事类型的调节作用显著，仅在个人道德两难故事中存在较小但稳定的效应；记分方式仅在非个人道德两难故事中的调节作用显著，只有多点记分存在较小但稳定的效应；性别和语系类型的调节作用均不显著。

参考文献

*元分析用到的参考文献

蔡阿燕, 杨洁敏, 许爽, 袁加锦. (2016). 表达抑制调节负性情绪的男性优势——来自事件相关电位的证据. *心理学报, 48*(05), 482–494.

陈武英, 卢家楣, 刘连启, 林文毅. (2014). 共情的性别差异. *心理科学进展, 22*(09), 1423–1434.

高园园. (2017). 道德判断中的外语效应及其心理机制. *教育观察, 6*(21), 142–144.

*关旭. (2017). *道德判断的外语效应：审慎加工被增强抑或直觉加工被弱化?* (硕士学位论文), 宁波大学.

关旭, 张锋, 周楠, 尹军. (2017). 决策行为和道德判断过程的外语效应：现状与展望. *应用心理学, 23*(01), 80–91.

*胡宏安. (2019). *道德外语效应的影响因素研究.* (硕士学位论文), 云南师范大学.

胡艺馨, 何英为, 王大伟. (2018). 道德决策中的情绪作用. *山东师范大学学报(人文社会科学版), 63*(06), 124–133.

黄龙, 徐富明, 胡笑羽. (2020). 眼动轨迹匹配法：一种研究决策过程的新方法. *心理科学进展, 28*(09), 1454–1461.

李明晖, 饶俪琳. (2017). 解释水平视角下的道德判断. *心理科学进展*(8), 1423–1430.

李笑燃, 肖前国, 吴小勇. (2014). 近十年有关道德的形成与加工机制研究述评. *内蒙古师范大学学报(教育科学版), 27*(12), 42–45.

林蓝菲, 高华. (2020). 道德判断的外语效应与框架效应. *心理技术与应用, 8*(11), 671–678.

罗跃嘉, 李万清, 彭家欣, 刘超. (2013). 道德判断的认知神经机制. *西南大学学报(社会科学版), 39*(03), 81–86.

倪传斌. (2020). 外语效应的表现形式、影响因素及其作用机制. *外语教学与研究, 52*(03), 397–408.

邱俊杰, 张锋. (2015). 道德困境中行为判断的认知与情绪问题:从道德双加工模型到建构水平理论. *应用心理学, 21*(003), 271–280.

田学红，杨群，张德玄，张烨. (2011). 道德直觉加工机制的理论构想. *心理科学进展, 19*(10), 1426–1433.

*杨雯雯. (2020). *道德判断中意图和结果信息加工的外语效应研究*(硕士学位论文). 电子科技大学，成都.

尹军，关旭，花蕊，张锋. (2018). 审慎加工对形成道德判断外语效应的作用. *应用心理学, 24*(04), 227–237.

余宏波，刘桂珍. (2006). 移情、道德推理、观点采择与亲社会行为关系的研究进展. *心理发展与教育, 22*(001), 113–116.

喻丰，彭凯平，韩婷婷，柴方圆，柏阳. (2011). 道德困境之困境——情与理的辩争. *心理科学进展, 19*(11), 1702–1712.

张诗容，胡平. (2017). 外语效应：证据，机制与前瞻. *中国临床心理学杂志, 25*(001), 45–49.

曾笑雨，马燚娜. (2020). 多项式模型在道德判断研究中的应用. *科学通报, 65*(19), 1912–1921.

张亚利，李森，俞国良. (2019). 自尊与社交焦虑的关系：基于中国学生群体的元分析. *心理科学进展, 027*(06), 1005–1018.

Aguilar, P., Brussino, S., & Fernández-Dols, J.-M. (2013). Psychological distance increases uncompromising consequentialism. *Journal of Experimental Social Psychology, 49*(3), 449–452. https://doi.org/10.1016/j.jesp.2013.01.002

Amit, E., Algom, D., & Trope, Y. (2009). Distance-dependent processing of pictures and words. *Journal of Experimental Psychology, 138*(3), 400–415. https://doi.org/10.1037/a0015835

Assink, M., & Wibbelink, C. J. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology, 12*(3), 154–174. https://doi.org/10.20982/tqmp.12.3.p154

Baron, J. (1994). Nonconsequentialist decisions. *Behavioral Brain Sciences, 17*(1), 1–10. https://doi.org/10.1017/S0140525X0003301X

Bialek, M., Paruzel-Czachura, M., & Gawronski, B. (2019). Foreign language effects on moral dilemma judgments: An analysis using the CNI model. *Journal of Experimental Social Psychology*, *85*, Article 103855. https://doi.org/10.1016/j.jesp.2019.103855

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (Eds.) (2011). *Introduction to meta-analysis*. John Wiley & Sons.

*Brouwer, S. (2019). The auditory foreign-language effect of moral decision making in highly proficient bilinguals. *Journal of Multilingual and Multicultural Development, 40*(10), 865–878. https://doi.org/10.1080/01434632.2019.1585863

Caldwell-Harris, C. L. (2014). Emotionality differences between a native and foreign language: Theoretical implications. *Frontiers in Psychology, 5*, 1055–1058. https://doi.org/10.3389/fpsyg.2014.01055

*Cavar, F., & Tytus, A. E. (2017). Moral judgement and foreign language effect: When the foreign language becomes the second language. *Journal of Multilingual and Multicultural Development, 39*(1), 17–28. https://doi.org/10.1080/01434632.2017.1304397

*Chan, Y. L., Gu, X., Ng, J. C. K., & Tse, C. S. (2016). Effects of dilemma type, language, and emotion arousal on utilitarian vs deontological choice to moral dilemmas in Chinese-English bilinguals. *Asian Journal of Social Psychology, 19*(1), 55–65. https://doi.org/10.1111/ajsp.12123

*Cipolletti, H., McFarlane, S., & Weissglass, C. (2016). The moral foreign-language effect. *Philosophical Psychology, 29*(1), 23–40. https://doi.org/10.1080/09515089.2014.993063

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Colomé, à., & Miozzo, M. (2010). Which words are activated during bilingual word production? *Journal of Experimental Psychology: Learning Memory & Cognition, 36*(1), 96–109. https://doi.org/10.1037/a0017677

Conrad, M., Recio, G., & Jacobs, A. M. (2011). The time course of emotion effects in first and second language processing: A cross cultural ERP study with German–Spanish bilinguals. *Frontiers in Psychology, 2*, 351–366. https://doi.org/10.3389/fpsyg.2011.00351

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology, 104*(2), 216–235. https://doi.org/10.1037/a0031021

*Corey, J. D., Hayakawa, S., Foucart, A., Aparici, M., Botella, J., Costa, A., & Keysar, B. (2017). Our moral choices are foreign to us. *Journal of Experimental Psychology: Learning Memory and Cognition, 43*(7), 1109–1128. https://doi.org/10.1037/xlm0000356

*Costa, A., Corey, J. D., Hayakawa, S., Aparici, M., Vives, M. L., & Keysar, B. (2017). The role of intentions and outcomes in the foreign language effect on moral judgements. *Quarterly Journal of Experimental Psychology, 72*(1), 8–17. https://doi.org/10.1177/1747021817738409

*Costa, A., Foucart, A., Hayakawa, S., Aparici, M., Apesteguia, J., Heafner, J., & Keysar, B. (2014). Your morals depend on language. *PloS one, 9*(4), Article e94842. https://doi.org/10.1371/journal.pone.0094842

Crain, WC (1985). *Theories of Development.* Prentice-Hall, New York, pp. 118-136

Cushman, F., Young, L., & Greene, J. D. (2010). Our mutli-system moral psychology: Towards a consensus view. In J. Doris, G. Harman, S. Nichols, J. Prinz, W. Sinnott-Armstrong, S. Stich (Eds.), *The Oxford handbook of moral psychology* (pp. 47–72). Oxford, U.K.

Degner, J., Doycheva, C., & Wentura, D. (2012). It matters how much you talk: On the automaticity of affective connotations of first and second language words. *Bilingualism: Language and Cognition, 15*(1), 181─189. https://doi.org/10.1017/S1366728911000095

Dewaele, J.-M. (2004). The emotional force of swearwords and taboo words in the speech of multilinguals. *Journal of Multilingual & Multicultural Development, 25*(2─3), 204─222. https://doi.org/10.1080/01434630408666529

Dewaele, J.-M. (2010). Christ fucking shit merde! Language preferences for swearing among maximally proficient multilinguals. *Sociolinguistic Studies, 4*(3), 595─614. https://doi.org/10.1558/sols.v4i3.595

*Driver, M. Y. (2020). Switching codes and shifting morals: how code-switching and emotion affect moral judgment. *International Journal of Bilingual Education and Bilingualism, 1005,* 1─17. https://doi.org/10.1080/13670050.2020.1730763

Dylman, A. S., & Barry, C. (2018). When having two names facilitates lexical selection: Similar results in the picture-word task from translation distractors in bilinguals and synonym distractors in monolinguals. *Cognition, 171*, 151─171. https://doi.org/10.1016/j.cognition.2017.09.014

*Dylman, A. S., & Champoux-Larsson, M.-F. (2020). It's (not) all Greek to me: Boundaries of the foreign language effect. *Cognition, 196*, Article 104148. https://doi.org/10.1016/j.cognition.2019.104148

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629─634. https://doi.org/10.1136/bmj.315.7109.629

Fisher, Z., & Tipton, E. (2015). Robumeta: An R-package for robust variance estimation in meta-analysis. *arXiv preprint arXiv:.02220*.

Friesdorf, R., Conway, P., & Gawronski, B. (2015). Gender differences in responses to moral dilemmas: A process dissociation analysis. *Personality Social Psychology Bulletin, 41*(5), 696─713. https://doi.org/10.1177/0146167215575731

Fu, R., Gartlehner, G., Grant, M., Shamliyan, T., Sedrakyan, A., Wilt, T. J., Ismaila, A. (2011). Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. *Journal of Clinical Epidemiology, 64*(11), 1187─1197. https://doi.org/10.1016/j.jclinepi.2010.08.010

*Geipel, J., Hadjichristidis, C., & Surian, L. (2015a). The foreign language effect on moral judgment: The role of emotions and norms. *PloS one, 10*(7), Article e0131529. https://doi.org/10.1371/journal.pone.0131529

*Geipel, J., Hadjichristidis, C., & Surian, L. (2015b). How foreign language shapes moral judgment. *Journal of Experimental Social Psychology, 59*, 8–17. https://doi.org/10.1016/j.jesp.2015.02.001

*Geipel, J., Hadjichristidis, C., & Surian, L. (2016). Foreign language affects the contribution of intentions and outcomes to moral judgment. *Cognition, 154*, 34–39. https://doi.org/10.1016/j.cognition.2016.05.010

Gong, H., & Medin, D. L. (2012). Construal levels and moral judgment: Some complications. *Judgment and Decision Making, 7*(5), 628–638.

Goulden, & Keith, J. (2006). Effect sizes for research: A broad practical approach. *Journal of Developmental Behavioral Pediatrics, 27*(5), 419–420.

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality & Social Psychology, 101*(2), 366–385. https://doi.org/10.1037/a0021847

Green, D. W. (2009). Mental control of the bilingual lexico-semantic system. *Bilingualism Language & Cognition, 1*(2), 67–81. https://doi.org/10.1017/S1366728998000133

Hadjichristidis, C., Geipel, J., & Savadori, L. (2015). The effect of foreign language in judgments of risk and benefit: The role of affect. *Journal of Experimental Psychology: Applied, 21*(2), 117–129. https://doi.org/10.1037/xap0000044

Hadjichristidis, C., Geipel, J., & Surian, L. (2017). How foreign language affects decisions: Rethinking the brain-drain model. *Journal of International Business Studies, 48*(5), 645–651. https://doi.org/10.1057/s41267-016-0040-1

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814–834. https://doi.org/10.1037/0033-295X.108.4.814

Haidt, J. (2007). The new synthesis in moral psychology. *Science, 316*(5827), 998–1002. https://doi.org/10.1126/science.1137651

Haidt, J. (2008). Morality. *Perspectives on Psychological Science, 3*(1), 65–72.

Harenski, C. L., Antonenko, O., Shane, M. S., & Kiehl, K. A. (2010). A functional imaging investigation of moral deliberation and moral intuition. *Neuroimage, 49*(3), 2707–2716. https://doi.org/10.1016/j.neuroimage.2009.10.062

Harris, C. L., Ayçíçeğí, A., & Gleason, J. B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Applied Psycholinguistics, 24*(4), 561–579.

Harris, C. L., Gleason, J. B., & Aycicegi, A. (2006). When is a first language more emotional? Psychophysiological evidence from bilingual speakers. *Bilingual Education Bilingualism, 56*, 257–283.

Hayakawa, S., Costa, A., Foucart, A., & Keysar, B. (2016). Using a foreign language changes our choices. *Trends in Cognitive Sciences, 20*(11), 791–793. https://doi.org/10.1016/j.tics.2016.08.004

Hayakawa, S., Tannenbaum, D., Costa, A., Corey, J. D., & Keysar, B. (2017). Thinking more or feeling less? Explaining the foreign-language effect on moral judgment. *Psychological Science, 28*(10), 1387–1397. https://doi.org/10.1177/0956797617720944

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Holman, E. W., Brown, C. H., Wichmann, S., Müller, A., Velupillai, V., Hammarström, H., Brown, P. (2011). Automated dating of the world's language families based on lexical similarity. *Current Anthropology, 52*(6), 841–875. https://doi.org/10.1086/662127

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*(9), 697–720. https://doi.org/10.1037/0003-066X.58.9.697

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics Biases: The Psychology of Intuitive Judgment, 49*, 81–111. https://doi.org/10.1017/CBO9780511808098.004

Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science, 23*(6), 661–668. https://doi.org/10.1177/0956797611432178

Kuppens, S., Laurent, L., Heyvaert, M., & Onghena, P. (2013). Associations between parental psychological control and relational aggression in children and adolescents: A multilevel and sequential meta-analysis. *Developmental Psychology, 49*(9), 1697–1712. https://doi.org/10.1037/a0030740

Liberman, N., & Trope, Y. (2014). Traversing psychological distance. *Trends in Cognitive Sciences, 18*(7), 364–369. https://doi.org/10.1016/j.tics.2014.03.001

*Miozzo, M., Navarrete, E., Ongis, M., Mello, E., Girotto, V., & Peressotti, F. (2020). Foreign language effect in decision-making: How foreign is it? *Cognition*, *199*, Article 104245. https://doi.org/10.1016/j.cognition.2020.104245

Moll, J., & de Oliveira-Souza, R. (2007). Moral judgments, emotions and the utilitarian brain. *Trends in Cognitive Sciences, 11*(8), 319–321. https://doi.org/10.1016/j.tics.2007.06.001

Moore, A. B., Lee, N. L., Clark, B. A., & Conway, A. R. (2011). In defense of the personal/impersonal distinction in moral psychology research: Cross-cultural validation of the dual process model of moral judgment. *Judgment Decision Making, 6*(3), 186–195. http://journal.sjdm.org/11/10310/jdm10310.html

*Muda, R., Niszczota, P., Bialek, M., & Conway, P. (2018). Reading dilemmas in a foreign language reduces both deontological and utilitarian response tendencies. *Journal of Experimental Psychology: Learning Memory and Cognition, 44*(2), 321–326. https://doi.org/10.1037/xlm0000447

Pavlenko, A. (2004). 'Stop doing that, Ia Komu Skazala!': Language choice and emotions in parent—child communication. *Journal of Multilingual and Multicultural Development, 25*(2-3), 179–203. https://doi.org/10.1080/01434630408666528

Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology, 47*(6), 405–428. https://doi.org/10.1080/00207594.2012.743665

Pereira, I. A. D., & Praxedes, P. H. L. (2016). Bilingualism, emotions and morality: A study based on the appraisal system. *Entrepalavras, 6*(2), 213–236.

Polonioli, A. (2018). A blind spot in research on foreign language effects in judgment and decision-making. *Frontiers in Psychology, 9*, 227–231. https://doi.org/10.3389/fpsyg.2018.00227

Puntoni, S., De Langhe, B., & Van Osselaer, S. M. (2009). Bilingualism and the emotional intensity of advertising language. *Journal of Consumer Research, 35*(6), 1012–1025. https://doi.org/10.1086/595022

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). Publication bias in meta-analysis. *Publication Bias in Meta-Analysis: Prevention, Assessment Adjustments*, 1–7. https://doi.org/10.1002/0470870168.ch1

Schwarzer, G. (2007). Meta: An R package for meta-analysis. *R news, 7*(3), 40–45.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*: Routledge.

*Shin, H. I., & Kim, J. (2017). Foreign language effect and psychological distance. *Journal of Psycholinguistic Research, 46*(6), 1339–1352. https://doi.org/10.1007/s10936-017-9498-7

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better P-curves: Making P-curve analysis more robust to errors, fraud, and ambitious P-hacking, a Reply to Ulrich and Miller (2015). *Journal of Experimental Psychology: General, 144(6)*, 1146–1152. https://doi.org/10.1037/xge0000104

Sinnott-Armstrong, W., Young, L., & Cushman, F. (2010). Moral intuitions as heuristics. *The Moral Psychology Handbook, 1*, 246–272. https://doi.org/10.1093/acprof:oso/9780199582143.003.0008

Smith, T. C., Spiegelhalter, D. J., & Thomas, A. (1995). Bayesian approaches to random-effects meta-analysis: A comparative study. *Statistics in Medicine, 14*(24), 2685–2699. https://doi.org/10.1002/sim.4780142408

Thacker, S. B. (1988). Meta-analysis: A quantitative approach to research integration. *Jama, 259*(11), 1685–1689. https://doi.org/10.1001/jama.1988.03720110047033

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*(2), 440–463. https://doi.org/10.1037/a0018963

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453–458. https://doi.org/10.1126/science.7455683Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Vøllestad, J., Nielsen, M. B., & Nielsen, G. H. (2012). Mindfulness- and acceptance-based interventions for anxiety disorders: A systematic review and meta-analysis. *British Journal of Clinical Psychology, 51*(3), 239–260. https://doi.org/10.1111/j.2044-8260.2011.02024.x

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., ... & Matzke, D. (2017). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35–57. https://doi.org/10.3758/s13423-017-1343-3

Warn, D., Thompson, S., & Spiegelhalter, D. (2002). Bayesian random effects meta-analysis of trials with binary outcomes: Methods for the absolute risk difference and relative risk scales. *Statistics in Medicine, 21*(11), 1601–1623. https://doi.org/10.1002/sim.1189

Wong, G., & Ng, B. C. (2018). Moral judgement in early bilinguals: Language dominance influences responses to moral dilemmas. *Frontiers in Psychology*, *9*, 1070–1079. https://doi.org/10.3389/fpsyg.2018.01070

*Woumans, E., Van der Cruyssen, I., & Duyck, W. (2020). Crime and punishment: Morality judgment in a foreign language. *Journal of Experimental Psychology: General*, *149*(8), 1597–1602. https://doi.org/10.1037/xge0000736

Moral foreign language effect and its moderating variables: a systematic review and meta-analysis

ZHU Lin[1], LIU Jinru[1], LI Jing[1], LIU Conghui[1]

([1] *Department of Psychology, Renmin University of China, Beijing* 100872, *China*)

Abstract: When presented with the moral decision-making situation within a foreign language context, the individuals will show a stronger utilitarian moral evaluation tendency compared with the same situation within the native language context. This phenomenon is defined as the moral foreign language effect. Numerous research studies have explored how the language context can influence moral judgment. However, the results were far from consistent. In this work, we used meta-analysis to explore the effect of the language type (native language vs. foreign language) on the individuals' utilitarian tendency in moral judgments, and we analyzed several moderating variables. We retrieved 19 papers from the literature, with 46 independent samples, 97 effect sizes and 9672 participants that met the inclusion criteria. The results demonstrated a small but stable moral foreign language effect (g = 0.23). Analyzing the moderation factors indicated that the moral foreign language effect was influenced by the story type, with a small but stable effect for personal moral dilemma stories (g = 0.32). On the other hand, there was no foreign language effect for impersonal moral dilemma stories (g = 0.11) or daily moral evaluation stories (g = 0.12). In addition, the foreign language effect for impersonal moral dilemma stories was affected by the scoring method. There was a significant effect under multi-point scoring (g = 0.27), but not under two-point scoring (g = 0.05). As for the sex and language family type, no significant moderating effect was observed. Our results suggest that the language type has a certain impact on the individuals' decision tendency when faced with moral dilemmas. More attention should be paid to the influence of the moral dilemma story type and a scoring method in future research.

Key words: moral judgment, foreign language effect, meta-analysis, moderating effect